FEMINIST DATA SET

CAROLINE SINDERS

ABOUT THE AUTHOR

Caroline Sinders is a machine-learning-design researcher and artist. For the past few years, she has been examining the intersections of technology's impact in society, interface design, artificial intelligence, abuse, and politics in digital, conversational spaces. Sinders is the founder of Convocation Design + Research, an agency focusing on the intersections of machine learning, user research, designing for public good, and solving difficult communication problems. As a designer and researcher, she has worked with Amnesty International, Intel, IBM Watson, the Wikimedia Foundation, and others.

Sinders has held fellowships with the Harvard Kennedy School, the Mozilla Foundation, Yerba Buena Center for the Arts, Eyebeam, STUDIO for Creative Inquiry, and the International Center of Photography. Her work has been supported by the Ford Foundation, Omidyar Network, the Open Technology Fund and the Knight Foundation. Her work has been featured in the Tate Exchange in Tate Modern, Victoria and Albert Museum, MoMA PS1, LABoral, Ars Electronica, the Houston Center for Contemporary Craft, *Slate*, *Quartz*, *Wired*, as well as others. Sinders holds a Masters from New York University's Interactive Telecommunications Program.

CLINIC FOR OPEN SOURCE ARTS
DU.EDU/AHSS/OPENSOURCEARTS

This project was sponsored in part by the Clinic for Open Source Arts (COSA) at the University of Denver.

The layout was designed by Kat Coneybear, a graphic designer living and working in New Orleans, Louisiana.

TABLE OF CONTENTS

INTRO	What is Feminist Data? And Why Do We Need A Feminist Data Set?	*4*
1	Making Critical Ethical Software	*6*
2	Thinking Through Feminist Data Collection and Creation	*12*
3	Thinking Through AI From Start to Finish	*TBD*
4	Ethical Design and Data Movements: Recognizing and Honoring Other Radical Critiques of Technology and Structures	*TBD*
5	Art as Policy	*TBD*
6	Feminist Data Set In Action	*TBD*
7	Analyzing Bias, Power and Privilege	*TBD*
8	Design Thinking to Find the Idea	*TBD*
9	Labor and AI: Training and Building AI Systems and Models	*TBD*
10	Can Data Models be Feminist?	*TBD*
11	Language Has Meaning: Researching How AI Interpreted Language	*TBD*
12	Your Homemade Guide to Interrogating Data at Home	*TBD*
13	What Does a Feminist Chat Bot Need? Thinking Through Conversational Design	*TBD*
14	Criticial User Interfaces: Using Feminist HCI and UI Design to Create	*TBD*

WHAT IS FEMINIST DATA? AND WHY DO WE NEED A FEMINIST DATA SET?

Feminist Data Set is a multi-year project that interrogates every step of the AI process that includes data collection, data labeling, data training, selecting an algorithm to use, the algorithmic model, and then designing how the model is then placed into a chat bot (and what the chatbot looks like) through intersectional feminism as an investigatory framework. Every step exists to question and analyze the pipeline of creating using machine learning—is each step feminist, is it intersectional, does each step have bias and how can that bias be removed? Really, what does it mean to think through every step slowly and thoughtfully; metaphorically, can we think of this as farm to server table, as slow data and consensual data?

Feminist Data Set is a public facing, social justice art practice. For example, the collecting of feminist data is held through public workshops and forums. The use of workshops and community input is inspired by the open source model of knowledge collection, Wikipedia. While Wikipedia is not without its faults, such as having an extremely white and male user base that can be riddled with toxicity.[1] However, the idea here stands that not one person or one single entity should be responsible for knowledge creation or gathering, it should fall to a community to reflect the community's ideas. After each workshop, as the creator, I then shift through submissions—to either track down provocations or ideas submitted by attendees or to verify submitted data to ensure that it's intersectional. Feminist

Feminist Data Set imagines data creation, as well as data sets and archiving, as an act of protest.

Data Set imagines data creation, as well as data sets and archiving, as an act of protest. When users, when people are cut out of the process of deciding what data collection means and how data is collected, the communal and slow process of Feminist Data Set empahsizes this protest. Feminist Data Set is a process-driven art project. In a time where so much personal data is caught and hidden by large technology companies and used for targeted advertising and algorithmic suggestions, what does it mean to make a data set about political ideology, one designed for use as protest, and to make a data set as a community? Feminist Data Set as a multi-year project aims to confront this. In our first workshop in 2017, over three days at SPACE art and technology in London, I and other participants started to conceptualize the greater project by writing our manifesto, featured at the bottom of this worksheet, and thinking through structures of data sets, structures of data collection, and how structure can

[1] *https://www.theatlantic.com/technology/archive/2015/10/how-wikipedia-is-hostile-to-women/411619/*

be inhibiting and also revolutionary. This idea of the structure inhibiting and/or rendering invisibility while also providing spaces for creation will be explored in a later chapter. In the data collection section, this duality on data will be explored—data as constriction and data as openness.

Politically and artisticially, Feminist Data Set is is inspired by the work of the maker movement, critical design, Arte Útil, Data Feminism, Design Justice, the Critical Engineering Manifesto, Xenofeminism, and the Feminist Principles of the Internet. Pedagogically, Feminist Data Set operates in a similar vein to Thomas Thwaites's "Toaster Project," a critical design project in which Thwaites builds a commercial toaster from scratch. Feminist Data Set, however, takes a critical and artistic view on software, particularly machine learning. What does it mean to thoughtfully make machine learning, to carefully consider every angle of making, iterating, and designing? Every step of this process needs to be thoroughly re-examined through a feminist lens.

Currently, the project has just explored data collection and data training and labeling. I've already described, at a high level, the data collection part of Feminist Data Set, and in another chapter, the data training and labeling will be described. In machine learning, once data has been collected, it needs to be labeled and trained for the data model for the algorithm, and this process needs to be analyzed as well. Data labeling and training are done by people, and often done on platforms like Amazon's Mechanical Turk. However, workers on these platforms are often severely underpaid. A platform that allows for pay inequity, doesn't utilize the governance structures they have in place to benefit workers,[2] and doesn't seem to check tasks for abuse isn't a feminist system. Through workshops, we've started exploring what an intersectionally feminist data modeling and training system would need. Domains that need to be addressed are pay equity, transparency for tasks, working governance systems to ensure safety for workers and ensure quality of tasks, a way to refute abuse or mislabeled tasks, communication between workers on the platform, and a reporting system on clients that can be shared between workers. Currently, for the Feminist Data Set project we've created an open source training tool and a wage calculator to help creators, artists, start ups and people better price tasks for training and labeling data sets to try to create an intervention and more transprency around wage inequity.

[2] *https://www.nytimes.com/interactive/2019/11/15/nyregion/amazon-mechanical-turk.html*

CHAPTER 1

MAKING CRITICAL ETHICAL SOFTWARE

(originally published for Dilettante Army in Winter 2018, updated for this reader in 2020)

Machine learning is a large and important part of product design. It's in most everyday things that users touch—from Google's search results, to internet ads, to Netflix recommendations. Machine learning algorithms have also been used in biased and unjust prison sentencing,[2] in facial detection[3] (or lack thereof), in biased hiring practices that favor male applicants[4] in technical roles, and in other ways that harm everyday people in everyday ways. Technology can amplify harm, bias, misogyny, racism, and white supremacy. Even if technology cannot create equality, technology as a whole needs to be examined and remade in order to create some spaces of equity.

What would it look like to create technology that acts as harm reduction, that acts actively as critique, as an artistic as well as a technical intervention? Where are the intersections of machine learning with tech and with art? Tania Bruguera's concept of Arte Útil, or "Useful Art," explores this notion of usefulness, of utilitarianism, to create an intervention that is both an artwork and a tool.

> *Useful Art is a way of working with aesthetic experiences that focus on the implementation of art in society where art's function is no longer to be a space for "signaling" problems, but the place from which to create the proposal and implementation of possible solutions. We should go back to the times when art was not something to look at in awe, but something to generate from. If it is political art, it deals with the consequences, if it deals with the consequences, I think it has to be useful art.*[5]

Feminist Data Set started in 2017 as a response, from an art and critical design space, to the many documented cases of problems in technology as well as bias in machine learning. It is a critical research and art project that examines bias in machine learning through data collection, data training, neural networks, and new forms of user interface (UI), as well as an attempt to create a feminist artificial intelligence (AI). Inspired by the work of the maker movement, critical design, Arte Útil, the critical engineering manifesto, Xenofeminism, and the Feminist Principles of the Internet, Feminist Data Set is a multi-year project designed to counteract bias in data and machine learning.

In machine learning, data is what defines the algorithm: it determines what the algorithm does. In this way, data is activated: it has a particular purpose, it is as important as the code of the algorithm. But so many algorithms exist as proprietary software, as black boxes that are impossible to unpack. For instance, why does Facebook's timeline serve up certain posts at certain times by certain users? Why do you see some friends' photos more than others? Facebook's timeline is driven by a black box algorithm designed to give you selective and "important" content; these standards of "importance" are determined by Facebook and are not shared with the general public, nor are they open to public critique or feedback. How Facebook responds and processes a user's data is considered proprietary; as a user, you do not own your data, Facebook does. How Facebook chooses to feed your own data back to you is considered Facebook's proprietary content.

[2] *Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner, 'Machine Bias', ProPublica, 23 May 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.*

[3] *Adam Frucci, 'HP Face-Tracking Webcams Don't Recognise Black People', Gizmodo, 22 December 2009, https://www.gizmodo.com.au/2009/12/hp-face-tracking-webcams-dont-recognise-black-people/.*

[4] *Jeffrey Dastin, 'Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women,' Reuters, 10 October 2018, https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.*

[5] *Arte Útil, https://www.arte-util.org/.*

Rachel Steinberg, image of the artist holding a workshop to create the data taxonomy and matrix for Feminist Data Set, 2018. Image reprinted with permission of SOHO20 gallery.

Data inside of software, and especially in social networks, comes from people. What someone likes, when they talk to friends, and how they use a platform is human data—it's not cold, mechanical, or benign. Data inside of social networks is intimate data, because conversations and social interactions, be they IRL or online, are varying forms of intimacy. How people interact with each other are what they like, and post, and dislike are "things" that are completely human; those "things" are also data.

FEMINISM AND TECHNOLOGY

Feminist Data Set imagines data creation, as well as data sets and archiving, as an act of protest. In a time where so much personal data is caught and hidden by large technology companies, used for targeted advertising and algorithmic suggestions, what does it mean to make a data set about political ideology, one designed for use as protest? How can data sets come from creative spaces, how can they be communal acts and works? What is a data set about a community that is made by that community? It can be a self-portrait, it can be protest, it can be a demand to be seen, it can be intervention or confrontation, or all of the above. It can be incredibly political. What about how a system then interprets that data? What if that system were also open to critique as well as community input?

Ethical, communal, "hackable" design and technology is a start towards an equitable future. It allows for community input, and for a community to drive or change a decision about a product, its technical capability, and its infrastructure. Feminist Principles of the Internet, a manifesto that addresses how to build feminist technology for the internet, builds upon this ethos by recognizing that in the process of making technical infrastructure, who it's built for has political ramifications. Feminist Principles of the Internet pushes open source technology and communities further by demanding space for marginalized groups and intent within technology, and it's this ethos in which Feminist Data Set exists.

Feminist Principles of the Internet, as well as theories like cyborg feminism and Xenofeminism, calls for a change in technology and how it functions, as well as

a change of leadership and ownership for that technology. The manifesto of Feminist Principles of the Internet demands a redefinition and re-purposing of technology and open source: 'Women and queer persons have the right to code, design, adapt and critically and sustainably use ICTs and reclaim technology as a platform for creativity and expression, as well as to challenge the cultures of sexism and discrimination in all spaces.'[6] Within this document, Feminist Principles of the Internet define "agency" as a necessary form of empowerment. 'We call on the need to build an ethics and politics of consent into the culture, design, policies and terms of service of internet platforms. Women's agency lies in their ability to make informed decisions on what aspects of their public or private lives to share online.'[7] Feminist Data Set exists within those realms of both technology and agency, as a critique on current machine learning infrastructure and practices, as well as a technical framework, critical methodology, and practice-based artwork attempting to address these issues.

CRITICAL DESIGN AS A MEDIUM

This kind of re-examining of what software can do, and should do, is similar to the creation of "Critical Design," design that addresses the limitations of product design. Feminist Data Set is inspired by this critical lens, and takes aspects of the critical design movement but applies it to machine learning. "Critical Design" as a term was coined in 1997 by Anthony Dunne, and it comes from a practice he developed with Fiona Raby when they were research fellows at the Royal College of Art. Critical Design as a practice 'is one among a growing number of approaches that aim to prevent and define interrogative, discursive, and experimental approaches in design practice and research.'[8] Critical design demands that design stop existing in terms of capitalist production, which is a function of product design, and push product design towards self-examination and cultural critique. Dunne and Raby highlight that by acknowledging that:

> *the design procession needs to mature and find ways of operating outside the tight constraints of servicing industry. At its worst, product design simply reinforces capitalist values. Design needs to see this for what it is, just one possibility, and to develop alternative roles for itself. It needs to establish an intellectual stance of its own, or the design profession is destined to lose all intellectual credibility and [be] viewed simply as an agent of capitalism.*[9]

By redefining how product design can be created, critical design also creates new ways for an audience to engage with design and understand all of the different facets of design in their daily lives.

Similarly, the maker movement helped redefine hardware engineering as a space of exploration and engagement by redefining "engineering." Engineering went from an expertise held by seasoned computer scientists to a method of investigation and construction that also welcomed everyday tinkerers and explorers. The growing ease and quickness of creating technology moved engineering into a space of play, and with that, into a space where anyone could create, changing what it meant to be "an engineer." With the advent of the small, affordable, adaptable and easy-to-use Arduinos (microcomputers), the debut of the company Adafruit, the creation of the fair/conference Maker Faire (a gathering for maker enthusiasts), and now with LittleBits (an educational company that creates small toys that teach children about electrical engineering, a pre-step to getting an Arduino), the revolution of the maker movement is similar to the citizen science movement—it's a technology-for-everyone revolution. When technology components are easy to engage with, able to be both used and remixed, they create a diversity of projects as well as a diversity of community. It's revolutionary because it creates new spaces but also allows for the examination of processes, like traditional computer science, that once existed in the "walled garden" of the academy. When institutionalized processes are opened up, and become open-source, they are re-adapted for an entire community. These processes allow for more input, and one could argue, more equity. But that's in an ideal world—there are still plenty of issues of racism and sexism, as well as gatekeeping, inside both the open source movement and the maker movement. That being said, the ability to augment, change, remix, or "fork" an experience, technology and code is what makes "making" an important movement in the modern world.

Feminist Data Set operates in a similar vein to Thomas Thwaites's Toaster Project, a critical design project

[6] *'Feminist Principles of the Internet v. 2.0', Feminist Principles of the Internet, last modified August 26, 2016, https://feministinternet.org/sites/default/files/Feminist_principles_of_the_internetv2-0.pdf.*

[7] *Feminist Principles of the Internet, page 4.*

[8] *Matt Malpass, Critical Design in Context: History, Theory, and Practices, London: Bloomsbury, 2016.*

[9] *Anthony Dunne and Fiona Raby, Design Noir: The Secret Life of Electronic Objects. Basel: August/Birkhäuser, 2001.*

CAROLINE SINDERS
CAROLINE SINDERS
The Analytical Engine would never be completed but it represents the conceptual dawn of the computer age. The four components of its design input, storage, processing and output- remain core components of all computers today, and the strikingly original notes Ada (Lovelace) prepared to explain this new kind of machine would presage the literature of computer science by nearly a decade.

Caroline Sinders, The Feminist Data Set installed at the Victoria and Albert Museum, 2018. Image courtesy of the Victoria and Albert Museum, London.

in which Thwaites builds a commercial toaster from scratch—from melting iron ore to building circuits and creating a new plastic toaster body mold. Feminist Data Set, however, takes a critical and artistic view on software and particularly on machine learning. What does it mean to thoughtfully make machine learning, to carefully consider every angle of making, iterating, and designing? Every step of this process needs to be thoroughly re-examined through a feminist lens, and like Thwaites's toaster, every step has to actually work.

THE PROCESS OF FEMINIST DATA SET

Originally, Feminist Data Set started as collaborative data set built through a series of workshops to address bias in artificial intelligence. The iterative workshops are key; by "slowly" gathering data in physical workshops, we allow a community to define feminism. But I also, workshop by workshop, examine what is in the data set so that I can course correct to address bias. I viewed this as a farm-to-table sustainable data set "growing." Are there too many works by cisgender women or white women in the data set? Then I need to address that in future or follow-up workshops by creating a call for non-cis women, for women of color, and for pieces of work by trans creators. In December 2018, we held a Feminist Data Set workshop in a queer bookstore to archive and add queer poets, writers, artists and community organizers' works to the data set. By design, the project will eventually conclude in creating a feminist AI system. However, there are many steps involved in the process:

1. **data collection**
2. **data structuring and data training**
3. **creating the data model**
4. **designing a specific algorithm to interpret the data**
5. **questioning whether a new algorithm needs to be created to be 'feminist' in its interpretation or understanding of the data and the models**
6. **prototyping the interface**
7. **refining**

Every step exists to question and analyze the pipeline of creating using machine learning—is each step feminist, is it intersectional, does each step have bias and how can that bias be removed?

As a way to "design out bias," I look to the The Critical Engineering Manifesto by Julian Oliver, Gordan Savičić, and Danja Vasiliev. The Critical Engineering Manifesto outlines ten principles as a guide to creating engineering projects, code, systems and ideals. Similar to critical design, it exists to examine the role that engineering and code plays in everyday life, as well as art and creative coding projects. Principles 2 and 7 address the role of shifting technology:

> 2. The Critical Engineer raises awareness that with each technological advance our techno-political literacy is challenged...
>
> 7. The Critical Engineer observes the space between the production and consumption of technology. Acting rapidly to changes in this space, the Critical Engineer serves to expose moments of imbalance and deception.[10]

Similarly, this manifesto (as well as others) helped serve as a basis for writing the original Feminist Data Set Manifesto, created in workshop 0 of the Feminist Data Set in London's SPACE Art and Technology in October 2017.

This manifesto defined the current and future intentions of the project. Feminist Data Set must be useful, it must disrupt and create new inputs for artificial intelligence, and must also be a project that focuses on intersectional feminism.

To date, the project has only been gathering data. The next step in the project will be addressing data training and data collection. In machine learning, when it comes to labeling data and creating a data model for an algorithm, groups will generally use Amazon's labor force, Mechanical Turk, to label data. Amazon created Mechanical Turk to solve their own machine learning problem of scale: they needed large data sets trained and labeled. Using Mechanical Turk in machine learning projects is standard in the field; it is used everywhere from technology companies to research groups to help label data. For the Feminist Data Set, the question is: Is the Mechanical Turk system feminist? Is using a platform of the gig economy ethical, is it feminist, is it intersectional? A system that creates competition amongst laborers, that discourages a union, that pays pennies per repetitive task, and that creates nameless and hidden labor is not ethical, nor is it feminist.

THE FEMINIST DATA SET MANIFESTO

OUR INITIAL INTENTION: to create a data set that provides a resource that can be used to train an AI to locate feminist and other intersectional ways of thinking across digital media distributed online.

OUR INTENTIONS, in Practice, over the course of two days, we created a data set that questions, examines, and explores themes of dominance. Inspired by the cyborg manifesto, our intention to add ambivalence, and to disrupt the unities of truth/information, mediated by algorithmic computation when it comes to expressing power structures in forms of domination, particularly in relationship to intersectional feminism.

OUR FUTURE INTENTIONS are to create inputs for an artificial intelligence to challenge dominance by engaging in new materials and engaging with others. We are building, collaboratively a collection.

Through collaboration, we created a new way to augment intelligence and augmented intelligence systems instead focusing on autonomous systems.

OUR MAIN TERMS: disrupt, dominance

MANIFESTO: we are creating a space/thing/data set/capsule/art to question dominance.

[10] *https://www.nbcnews.com/think/opinion/google-search-algorithms-are-not-impartial-they-are-biased-just-ncna849886*

In 2020, I built, prototyped, and imagined what an ethical mechanical turk system could look like, one created from an intersectional feminist lens that can be used by research groups, individuals, and maybe even companies. Prior to building the system, I imagined and outlined the following in 2018:

The system should be ethical in the sense that it will allow for more transparency in who trains and labels a data set.

Could the trainers also be authors, and thus, the system will save data about the data set and its authors.

This system will also give researchers or project creators the ability to see how much one trainer has trained the data, as well as invite and verify new trainers.

Additionally, project creators will be able to pay trainers through this system by suggesting living wage payments.

This system also creates necessary data about the data set itself.

This data about the data set includes who labeled or trained it, where are they from, when the data set was "made" or finished, and what's in the data set (e.g. certain kinds of faces, etc.).

If machine learning is going to move forward in terms of transparency and ethics, then how data is trained, how trainers interact with it, and how datasets are used in algorithms and model creation need to be critically examined as well. Making must be thoughtful and critical to create equity. It must be open to feedback and interpretation. We must understand the role of data creation and how systems can use, misuse, and benefit from data. Data must be seen as something created from communities, and as a reflection of that community—data ownership is key. Data's position inside technology systems is political, it's activated, and it's intimate. For there to be equity in machine learning, every aspect of the system needs to be examined, taken apart and put back together. It needs to integrate the realities, contexts, and constraints of all different kinds of people, not just the ones who built the early Web. Technology needs to reflect those who are on the web now.

ACKNOWLEDGMENTS

This project would not be possible without the support of Rachel Steinberg, SOHO20 Gallery, the Victoria and Albert Museum, Sara Cluggage, and Irini Papadimitriou. This project exists in the same community, and thus could not exist without the following projects and creators: the Xenofeminism Manifesto, the works of Donna Harway, the Cyberfeminism Index, the Library of Missing Data Sets, the Feminist Principles of the Internet, and Kimberlé Crenshaw's canonical research.

REFERENCE LIST

Angwin, Julia, Jeff Larson, Surya Mattu and Lauren Kirchner. 'Machine Bias', ProPublica, 23 May 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Arte Útil, http://www.arte-util.org/.

The Critical Engineering Working Group. 'The Critical Engineering Manifesto', October 2011-2019, https://criticalengineering.org/.

Dastin, Jeffrey. 'Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women,' Reuters, 10 October 2018, https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G.

Dunne, Anthony and Fiona Raby. Design Noir: The Secret Life of Electronic Objects, Basel: August/Birkhäuser, 2001.

'Feminist Principles of the Internet v. 2.0', Feminist Principles of the Internet, last modified August 26, 2016, https://feministinternet.org/sites/default/files/Feminist_principles_of_the_internetv2-0.pdf.

Frucci, Adam. 'HP Face-Tracking Webcams Don't Recognise Black People', Gizmodo, 22 December 2009, https://www.gizmodo.com.au/2009/12/hp-face-tracking-webcams-dont-recognise-black-people/.

Malpass, Matt. Critical Design in Context: History, Theory, and Practices. London: Bloomsbury, 2016.

CHAPTER 2

THINKING THROUGH FEMINIST DATA COLLECTION AND CREATION:

A FEMINIST DATA SET DESIGN THINKING EXERCISE

This chapter and exerices are designed to help think through what is equitable and feminist data set collection, generally for machine learning, but these principles can be applied to most fields that use data. Below are descriptions and research on feminist data, as well as a few exercises to conduct to better understand bias in data sets, and what 'feminist data' in practice could be.

TERMS

data: individual facts, statistics, or items of information.[11]

data sets: a collection of data; usually, the data are related to each other or similar in some way. A data set could be all of your tweets you've ever tweeted, the interactions on a piece of digital content (for example, the amount of likes or shares on a Facebook post), or a data set of who lives in a neighborhood (like a Census), or how a city plants trees over time.

data collection: the process of gathering and measuring information or data on specific topics and variables, which are often related to a system and have a structure. Data collection can allow someone to potentially research and answer questions, and come to different outcomes. Data collection can be weaponized to harm by creating biased data sets, but can also be used to create archives and snapshots of information, places, and systems in different points of time.

machine learning: a system and a form of artificial intelligence that uses data and data sets to create outcomes and predictions from that data.

intersectional feminism: a form of feminism that acknowledges that different identities and marginalizations of a person should be viewed together as overlapping and not as separate issues. For example, looking at both race and gender in terms of marginalization that a woman of color would face.

INTERSECTIONAL FEMINISM AS FRAMEWORK

The goal of this exercise is to explore intersectional feminist thinking in data, but also in society. Who

[11] *Definition from the People's Guide to Ai, Onuoha and Nucera, page 8.*

Caroline Sinders leading a Feminist Data Set workshop in Berlin, DE for Z2X gathering.

technology is designed for, and how it acknowledges harms, pains, and pleasures, can have political implications. Intersectionality is a form of study that acknowledges that people have many different dimensions to their identities, and those dimensions can be added to marginalizations in society. Intersectional feminism, a term coined by Professor Kimberlé Crenshaw, is the acknowledgment that these marginalizations, and the addition of marginalization identities and aspects of identity, can lead to different experiences for groups of people, and thus, these differences need to be taken into account. For example, separating race from gender in the case of a black woman (as Crenshaw wrote in a law paper in 1991, where the term intersectional feminism was invented). Intersectional feminism acknowledges that the experiences of a black woman would differ from a white woman or a man of color, and that race and gender must be viewed as intersecting and overlapping identities. Intersectional feminism asks that we confront our biases and differences, and acknowledge the nuances of identity and privilege when investigating systems.

THINKING THROUGH DATA

Start with data collection. Let's revisit our definition of 'data collection.'

But what does this have to do with data? We need to think about power, context, and privilege in society, and how power, context, and privilege can change and warp technology, and produce unintended consequences. What is a feminist data set? Feminist Data Set is an intersectional, trans-inclusive, feminist art project that is collecting written work to train an artificial intelligence system to create a chatbot. Written work or 'text' is a form of data, just as census information and inputs is data, or images, like groups of dogs, muffins, or bananas, or the selfies of one person, are data. This 'text' that is being collected in Feminist Data Set through workshops can be anything text-based—from poetry, to blog posts, social media posts, podcasts transcripts, academic essays, books, articles, or music lyrics— that is written from a place of intersectional feminism. Similarly, feminist data would be data sets (of images, video, text, audio, you name it) that are created from an intersectional feminist point of view. Creating intersectional feminist data doesn't have to be data 'about feminism' or using the word feminism,

Image by Caroline Sinders from a Feminist Data Set workshop, 2017.

but it can be from a feminist point of view. An example of intersectional data could be an article that is on wage inequality. An intersectional feminist article would go beyond a binary wage inequality (of cis men being paid more than cis women) and acknowledge cis women of different races also face varying pay discrimination. An intersectional feminist article would outline what different women of different races make, since white women are paid more than black women, latinix/hispanic women, asian women and indigenous women.

But to do this, we need to first think about what we want to explore in the Feminist Data Set, and how can we find the 'right' topic? We can use design thinking exercises to start exploring a topic.

FINDING THE DATA

There is a bit of a reveal here when interacting the above exercise of 'finding' the data. The reveal is that written texts by femmes, cis women, non-gender binary folx, women and people of color, can be hard to find, especially if those texts, books, or articles are written from a feminist perspective. Marginalized groups and women are under-cited in scholarly articles, books, and press.[12] When holding Feminist Data Set workshops, even if there is access to a library, it's still hard to find this 'data,' the written texts. The searching for this kind of information—of articles written from an intersectional perspective—requires deeper and longer searching; it takes time. Even in multi-day workshops, it takes participants time to unpack, search, and search again for feminist data.

The difficulty beyond lack of citations extends to the design of search tools. The easiest to use and most accessible tool for participants is a web browser search engine, such as Google. But how search engines are designed poses a problem, especially in retrieving and ranking content. These search engines are biased[13] in surfacing content—some of it is from surfacing popular content,[14] and some of it is from misaligned

[12] *https://www.insidehighered.com/news/2018/08/16/new-research-shows-extent-gender-gap-citations*
[13] *https://www.technologyreview.com/2018/02/26/3299/meet-the-woman-who-searches-out-search-engines-bias-against-women-and-minorities/*
[14] *https://www.nbcnews.com/think/opinion/google-search-algorithms-are-not-impartial-they-are-biased-just-ncna849886*

or improperly trained algorithms. But the popular content feeding the algorithm for results is an interesting one—if marginalized groups are under-cited and their work is underreported, how can that content ever surface up high enough in Google search results? Search results help shape our access to information and content, and search engines create an ouroboros of popularized content, making it hard for new content to permeate through this loop. Feminist Data Set workshops highlight this problem through previously held workshops.

DIGGING DEEPER AND ORGANIZING OUR THOUGHTS

Materials needed: sticky notes, pens or pencils, paper

This is an exercise I run in the beginning of the Feminist Data Set workshop to help get participants thinking deeper about topics they will be exploring and researching. This works as an icebreaker or first activity because it's quiet, iterative, and really fast.

Set a timer for two minutes and start writing down any ideas or topics that come to mind. Try to use as few words as possible—not full sentences, just a pair of words to describe an idea. What are you interested in in exploring? Is it wage inequality, the pink tax, transgender rights, the history of indigenous activists, misogyny in music, copyright, citations and women scholars? Really, it can be anything. What's important to note is that we'll be doing this exercise multiple times, and you should view this as a quick way to get some ideas out there. Not all ideas will be your final ideas. But we are viewing this as a way to cast a net, and then whittle down to the ideas we want to explore.

So start writing down any ideas that come to mind. When the timer goes off, spend one minute looking at what you've collected. Pick your top two ideas. If you're in a group setting and feel comfortable, everyone could go around the room and read those top two ideas.

Now set a timer, and repeat the exercise, but try to go deeper on your ideas. Let's say you selected 'trans rights.' That's a big topic area—what specifically are you interested in? Is it how trans women are discriminated against? If so, what are the different kinds of discrimination? Are you thinking of an experience you've had or a friend has had or a story you've read? Try to write that down. Another idea may be pay inequality—what interests you about that? Is there a specific slice or sub topic? If you're generally interested in a broad topic (like pay discrimination, like the pink tax, like trans rights, bias in technology), try to write down all of the smaller and more specific ideas or topics within that broad topic.

Repeat this exercise up to three or four times.

Afterwards, take your sticky notes and start to organize them by theme. How are they related? If you're not sure yet, just start making small piles. After making these groupings, try naming them. This will help you organize your thoughts to then start searching for data. The more cohesive or concise your thoughts, the easier it is to find text on that topic like a blog post, an interview, an article, or archive.

Recommended reading:
https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=1052&context=uclf